|  |  |
| --- | --- |
| **Joint Collaborative Team on Video Coding (JCT-VC)**  **of ITU-T SG 16 WP 3 and ISO/IEC JTC 1/SC 29/WG 11**  33rd Meeting: Macao, CN, 6–12 Oct. 2018 | Document: JCTVC-AG0028-v2 |

|  |  |  |  |
| --- | --- | --- | --- |
| *Title:* | **HEVC encoder solution for composite long-term reference picture** | | |
| *Status:* | Input document to JCT-VC | | |
| *Purpose:* | Proposal | | |
| *Author(s) or Contact(s):* | Xiaozhen Zheng | Tel: Email: | xiaozhen.zheng@dji.com |
| *Source:* | SZ DJI Technology Co., Ltd., | | |

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

# Abstract

This contribution provide a HEVC encoder solution for composite long-term reference that has been evaluated in JVET CE11 at the meeting cycle of April 2018 to July 2018 [1]. Implementation details and test results followed by common test condition are provided in the document. Simulations show that the proposed technique can achieve -2.54%/-3.74%/-3.21% coding gain for Y/U/V over HM16.16 at Lowdelay B Main10 configuration (LDB) with around 18% encoding time increase.

# Introduction

Composite reference picture was proposed in JVET Call for Proposal [2] and it was studied in JVET CE11 at the meeting cycle of April 2018 to July 2018. As an alternative approach, HEVC encoder only implementation was evaluated at JVET CE11 test 3[1] and adopted to VTM and BMS software at Ljubljana JVET meeting [3].

This proposal provides an HM encoder only composite long-term reference implementation which is ported from VTM software and reports the test results. The basic generation process of composite long-term reference picture and its encoding workflow and the use of RPS syntax by exploiting HEVC mechanism are described in the document.

# Proposed method

## General idea

Composite long-term reference generates and updates a long-term reference by using information of static areas, which are always regarded as background, among a video scene. Usually, background areas have few motion in a long temporal window. Therefore the blocks with minor difference between the background and current frame are picked up, and are used to replace the co-located blocks in the long-term reference that targets at the renewal of the background information.

The proposed design targets at evaluating the use of the HEVC long-term reference mechanism, potentially in combination with signaling of no output coded pictures (pic\_output\_flag = 0). This combination could theoretically achieve similar functionality as that provided by composite reference pictures which is described in JVET CE11[1], e.g. by synthesizing and signaling a no-output reference picture that only contains background information which is illustrated at Fig. 1.



Fig. 1: *Example use of no output reference pictures*

As illustrated in Fig. 1, the frames I0, B1, B2, B3 and B4 are coded as short-term reference. During the encoding of those short-term frames, an alternative frame is composed by exploiting previous coded frames’ content. When an encoder determines the alternative frame has been completely constructed, such frame will be coded as long-term reference with pic\_output\_flag=0.

## Implementation details

### Modification of RPS setting in cfg file

In order to keep dpb process and the motion vector scaling process same as before when a long-term reference is encoded, all of short-term frames’ POC is doubled from “x” to “2x”. The RPS setting for short-term frames without long-term reference and short-term frames with long-term reference are showed at Fig. 2 and Fig. 3.

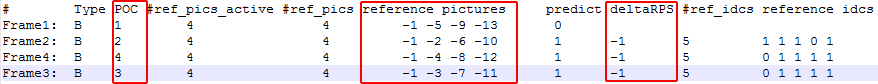


Fig2. RPS setting for short-term frames without long-term reference

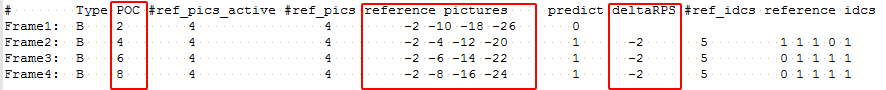


Fig3. RPS setting for short-term frames with long-term reference

### RPS syntax for long-term reference

A composed long-term reference is always coded after the first short-term frame and coded before the first frame of a particular GOP. Therefore, the parameter at SPS *long\_term\_ref\_pics\_present\_flag* is set to 1 and *num\_long\_term\_ref\_pics\_sps* is set 0.

The parameter at the composed long-term frame’s PPS *Output\_flag\_present\_flag* is set to 1, *deblocking\_filter\_control\_present\_flag* and *pps\_deblocking\_filter\_disabled\_flag* is set to 1.

Since RPS for the composed long-term frame is different to RPS at SPS, *short\_term\_ref\_pic\_set\_sps\_flag* is set to 0 at the composed long-term’s slice header.

Meanwhile, the reference pictures for the composed long-term frame are same as the reference pictures for the first frame in a particular coding GOP. Thus, *delta POC* for composed reference frame is set to -1, -9, -17, -25, *deltaRPS* is set to 1, *ref\_idcs* is set to 5 and *reference idcs* is set to 1, 1, 1, 1, 0. Fig.4 show an example for POC and deltaPOC setting when short-term frames are coded without composed long-term frame and with composed long-term frame.



Fig. 4: Illustration of short-term and long-term reference’s references

When a composed long-term reference is coded, a new short-term RPS need to be encoded with *inter\_ref\_pic\_set\_prediction\_flag* is set to 1, *delta\_rps*, *used\_by\_curr\_pic\_flag[j]* and *use\_delta\_flag[j]* are set to 1.

After the coded of the composed long-term frame, *num\_long\_term\_pics*, *used\_by\_curr\_pic\_lt\_flag* and *num\_ref\_idx\_active\_override\_flag* in the following short-term frames’ RPS is set to 1. *num\_ref\_idx\_l0\_active\_minus1* is set to 4.

## CTU selection strategy for composite reference update at encoder side

According to the principle of composite reference picture, CTU-level similarity metric between current original frame and two previous original frames is calculated before the coding of current frame. Based on CTU-level similarity metric, a certain number of CTUs of current frame are used and marked to updated composed reference. A control parameter, which is called as refresh rate and denoted as *R*, is used at encoder side which imposes the limitation on the total amount of CTUs in current frame that can be used to update composed reference. The value of *R* is set to 25% CTU number in a coding frame.

The CTU selection strategy is described below.

Denoted current frame’s coding order as N. The original frames with coding order N-1 and N-2 are denoted as img\_ref and img\_ref2 respectively. For each CTU, the following parameters are calculated:

DistY: the sum up of absolute difference between CTU in current frame and co-located CTU in img\_ref in Y component;

DistU: the sum up of absolute difference between CTU in current frame and co-located CTU in img\_ref in U component;

DistV: the sum up of absolute difference between CTU in current frame and co-located CTU in img\_ref in V component;

DistYRef2: the sum up of absolute difference between CTU in current frame and co-located CTU in img\_ref2 in Y component;

DistURef2: the sum up of absolute difference between CTU in current frame and co-located CTU in img\_ref2 in U component;

DistVRef2: the sum up of absolute difference between CTU in current frame and co-located CTU in img\_ref2 in V component;

LargeDist: the number of Luma pixel whose absolute difference value between pixel in current CTU and co-located CTU located at img\_ref or img\_ref2 is larger than 200<<(bitdepth - 8)

PixCount: Number of Luma pixels in current CTU

If either following conditionA or conditionB is true, the current CTU is marked as a candidate to update composed long-term reference:

* conditionA: all of the following conditions are true

- LargeDist / PixCount < 0.01

- DistY / PixCount < (3.5 << (bitdepth - 8))

- DistU / PixCount < (0.5 << (bitdepth - 8))

- DistV / PixCount < (0.5 << (bitdepth - 8))

* conditionB: all of the following conditions are true

- LargeDist / PixCount < 0.02

- DistY / PixCount < (6 << (bitdepth - 8))

- DistU / PixCount < (0.5 << (bitdepth - 8))

- DistV / PixCount < (0.5 << (bitdepth - 8))

- DistYRef2 / PixCount < (6 << (bitdepth - 8))

- DistURef2 / PixCount < (0.5 << (bitdepth - 8))

- DistVRef2 / PixCount < (0.5 << (bitdepth - 8))

After pick-up all candidate CTU, order all the candidates by using followed cost function:

cost = DistY / PixCount + DistU / PixCount + DistV / PixCount

Choose up to *R* CTUs with smallest cost as updated CTUs, and use their reconstruction pixels to replace the co-located blocks in composed long-term reference.

# Experimental results

The proposed composite long-term reference is implemented on HM16.16. The test is evaluated using BD rates, encoding time, decoding time according to Common Test Conditions (CTC) in [4].

## Simulation environment

For the simulations, same platform machines as below in PC cluster were used.

* Encode
  + OS: Windows Sever 2008 R2 Enterprise
  + Compiler: Visual Studio 2017
  + CPU: Intel Xeon E5-2687W v4 3.0GHz
    - SIMD: AVX2 instruction is enabled.
  + Memory: 256 GB
* Decode
  + OS: Windows Sever 2008 R2 Enterprise
  + Compiler: Visual Studio 2017
  + CPU: Intel Xeon E5-2687W v4 3.0GHz
    - SIMD: AVX2 instruction is enabled.
  + Memory: 256 GB

## Test results

The proposed HM encoder only solution’s test results are summarized in Table 1.

Table 1 : Over HM16.16 (LDB)

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | **Over HM16.16** | | | | |
|  | Y | U | V | EncT | DecT |
| Class A1 |  |  |  |  |  |
| Class A2 |  |  |  |  |  |
| Class B | -1.67% | -3.02% | -2.81% | 122% | 106% |
| Class C | -3.88% | -5.95% | -4.57% | 115% | 100% |
| Class E | -2.18% | -1.98% | -2.04% | 117% | 99% |
| **Overall** | -2.54% | -3.74% | -3.21% | 118% | 102% |
| Class D | 0.46% | 0.60% | 0.36% | 113% | 103% |

It should be noted that the proposed method need to generate a composite long-term reference from previous frames and need to be encoded in the bitstream. Therefore, the proposed method is not suitable for random access configuration since this configuration need to insert RAP every second roughly, which means that the long-term frame need to be encoded at every RAP and impose negative impact on coding performance.

Moreover, the encoding time of the proposed composite long-term reference can be speed up by shrinking motion search window for the long-term reference, which has been implemented on VTM 2.1 software.

# Conclusion

This contribution reports simulation result of composite long-term reference that implements HEVC encoder solution for composite reference. The average coding gain for the test is -2.54%/-3.74%/-3.21% for Y/U/V components over HM16.16 with about 18% encoding time increased. The proposed method shows substantial coding performance benefit for the scenario with lager background area. It’s suggested to adopt the proposed HEVC encoder solution for composite reference to the next version of HM software.

# References

[1] X. Zheng, G. Li and Y. Li, “Description of CE11: Composite reference picture” JVET-J1031, JVET 10th Meeting: San Diego, US, April 2018.

[2] Z. Wang, X. Meng, C. Jia, J. Cui, S. H. Wang, S. S. Wang, S. Ma, W. Li, Z. Miao, X. Zheng, “Description of SDR and HDR video coding technology proposal by DJI and Peking University” JVET-J0011, JVET 10th Meeting: San Diego, US, April 2018.

[3] G. J. Sullivan, J.-R. Ohm, “Meeting Report of the 11th meeting of the Joint Video Experts Team (JVET),

Ljubljana, SI, 10–18 July 2018”, JVET-K1000, 11th Meeting: Ljubljana, SI, 10–18 July 2018

[4] Jill Boyce, K. Suehring, X. Li, and Vadim Seregin, “JVET common test conditions and software reference configurations” JVET-J1010, April 2018.

# Patent rights declaration(s)

**SZ DJI Technology Co., Ltd. may have current or pending patent rights relating to the technology described in this contribution and, conditioned on reciprocity, is prepared to grant licenses under reasonable and non-discriminatory terms as necessary for implementation of the resulting ITU-T Recommendation | ISO/IEC International Standard (per box 2 of the ITU-T/ITU-R/ISO/IEC patent statement and licensing declaration form).**